

Breaking the Bonds of Submodularity

Empirical Estimation of Approximation Ratios for Monotone Non-Submodular Greedy Maximization

J. David Smith
CISE Department
University of Florida
Gainesville, Florida 32611
jdsmith@cise.ufl.edu

My T. Thai
CISE Department
University of Florida
Gainesville, Florida 32611
mythai@cise.ufl.edu

ABSTRACT

While greedy algorithms have long been observed to perform well on a wide variety of problems, up to now approximation ratios have only been known for their application to problems having *submodular* objective functions f . Since many practical problems have non-submodular f , there is a critical need to devise new techniques to bound the performance of greedy algorithms in the case of non-submodularity.

Our primary contribution is the introduction of a novel technique for estimating the approximation ratio of the greedy algorithm for maximization of monotone non-decreasing functions based on the curvature of f without relying on the submodularity constraint. We show that this technique reduces to the classical $(1 - 1/e)$ ratio for submodular functions. Furthermore, we develop an extension of this ratio to the adaptive greedy algorithm, which allows applications to non-submodular stochastic maximization problems. This notably extends support to applications modeling incomplete data with uncertainty. Finally we use this new technique to derive a $(1 - 1/\sqrt{e})$ ratio for a popular problem, Robust Influence Maximization [10], which is non-submodular and $1/2$ for Adaptive Max-Crawling [15], which is adaptive non-submodular.

ACM Reference format:

J. David Smith and My T. Thai. 2016. Breaking the Bonds of Submodularity. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17)*, 11 pages.
DOI: 10.1145/nnnnnnnn.nnnnnnnn

1 INTRODUCTION

It is well-known that greedy approximation algorithms perform remarkably well, especially when the traditional ratio of $(1 - 1/e) \approx 0.63$ [18] for maximization of *submodular* objective functions is considered. Over the four decades since the proof of this ratio, the use of greedy approximations has become widespread due to several factors. First, many interesting problems satisfy the property of *submodularity*, which states that the marginal gain of an element never increases. If this condition is satisfied, and the set of possible solutions can be phrased as a uniform matroid, then one of the highest general-purpose approximation ratios is available “for free”

with the use of the greedy algorithm. Second, the greedy algorithm is exceptionally simple both to understand and to implement.

A concrete example of this is the *Influence Maximization* problem, to which the greedy algorithm was applied with great success – ultimately leading to an empirical demonstration that it performed near-optimally on real-world data [16]. Kempe et al. showed this problem to be submodular under a broad class of influence diffusion models known as *Triggering Models* [11]. This led to a number of techniques being developed to improve the efficiency of the sampling needed to construct the problem instance (see e.g. [1, 19, 24] and references therein) while maintaining a $(1 - 1/e - \epsilon)$ ratio as a result of the greedy algorithm. This line of work ultimately led to a $(1 - \epsilon)$ -approximation by taking advantage the dramatic advances in sampling efficiency to construct an IP that can be solved in reasonable time [16]. In testing this method, it was found that greedy solutions performed near-optimally – an unexpected result given the $1 - 1/e$ worst-case.

For non-submodular problems, no approximation ratio for greedy algorithms is known. However, due to their simplicity they frequently see use as simple baselines for comparison. On the Robust Influence Maximization problem proposed by He & Kempe, the simple greedy method was used in this manner [10]. This problem consists of a non-submodular combination of Influence Maximization sub-problems and aims to address uncertainty in the diffusion model. Yet despite the non-submodularity of the problem, the greedy algorithm performed no worse than the bi-criteria approximation [10].

Another recent example of this phenomena is the socialbot reconnaissance attack studied by Li et al. [15]. They consider a minimization problem that seeks to answer how long a bot must operate to extract a certain level of sensitive information, and find that the objective function is (adaptive) submodular only in a scenario where users disregard network topology. In this scenario, the corresponding maximization problem, Max-Crawling, has a $1 - 1/e$ ratio due to the work of Golovin & Krause [8]. However, this constraint does not align with observed user behaviors. They give a model based on the work of Boshmaf et al. [2], who observed that the number of mutual friends with the bot strongly correlates with friending acceptance rate. Although this model is no longer adaptive submodular, the greedy algorithm still exhibited excellent performance. Thus we see that while submodularity is *sufficient* to imply good performance, it is not *necessary* for the greedy algorithm to perform well.

This, in turn, leads us to ask: is there any tool to theoretically bound the performance of greedy maximization with non-submodularity? Unfortunately, this condition has seen little study.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Conference'17, Washington, DC, USA

© 2016 Copyright held by the owner/author(s). 978-x-xxxx-xxxx-x/YY/MM...\$15.00
DOI: 10.1145/nnnnnnnn.nnnnnnnn

Wang et al. give a ratio for it in terms of the worst-case rate of change in marginal gain (the *elemental curvature* α) [26]. This suffices to construct bounds for non-submodular greedy maximization, though for non-trivial problem sizes they quickly approach 0. We note, however, that the α ratio still encodes strong assumptions about the worst case: that the global maximum rate of change can occur an arbitrary number of times.

Motivated by the unlikelihood of this scenario, our proposed bound instead works with an estimate of how much change can occur during the k steps taken by the greedy algorithm. In our experiments, we observe that in each case the total change converges towards a fixed value. This leads to new bounds for each of the above problems. We obtain a ratio of $1 - 1/\sqrt{e}$ for Robust Influence Maximization, providing an alternative to the $(1 - 1/e)OPT - \gamma$ ratio that requires an additional $1 + \ln|\Sigma| + \ln \frac{3}{\gamma}$ elements to be selected. Note that the value of OPT is on the range $[0, 1]$, and therefore γ must be small to obtain a competitive ratio. Further, we find that the Max-Crawling problem has a ratio of $1/2$ under the model matching observed user behavior. Finally, while we note little improvement over $1 - 1/e - \epsilon$ for the Influence Maximization problem, we observe that if a distribution over the number of differences between the optimal and greedy solutions can be estimated, significant improvements to the ratio can be made.

The remainder of this paper is arranged as follows: First, we briefly cover the preliminary material needed for the proofs and define the class of problems to which they apply (Sec. 1.1). We next define the notion of curvature used and develop a proof of the ratio based on it, with an extension to adaptive greedy algorithms, and show it is equivalent to the traditional $1 - 1/e$ ratio for submodular objectives (Sec. 2). This is followed by an algorithm to estimate the curvature of an objective in order to construct estimates from practical problem instances (Sec. 3). Then, we evaluate the ratio on a submodular problem instance with a known optimal solution to verify both that it acts as a lower bound and that it is of comparable quality to the traditional ratio (Sec. 4). Finally, we show our estimated $1 - 1/\sqrt{e}$ (Sec. 4.2) and $1/2$ (Sec. 4.3) ratios, each of which holds with probability $\delta = 0.95$, and conclude with a reflection on the contributions and a discussion of future work (Sec. 5).

Contributions.

- A technique for estimating the approximation ratio of greedy maximization of non-submodular monotone non-decreasing objectives on uniform matroids.
- An extension of this technique to adaptive greedy optimization, where future greedy steps depend on the success or failure of prior steps.
- An approximation ratio of $1 - 1/\sqrt{e}$ with 95% probability for greedy maximization of the non-submodular Robust Influence Maximization problem.
- An approximation ratio of $1/2$ with 95% probability for greedy maximization of the adaptive, non-submodular Max-Crawling problem.

1.1 Background & Related Work

To understand both the state of the art and advancements of this work, we first briefly cover each constraint required by the classical $1 - 1/e$ ratio [18].

1.1.1 Constraints on the $1 - 1/e$ Ratio.

Uniform Matroids. A matroid defines the notion of dependencies between elements of a set, and are denoted by $\mathcal{M} = (X, \mathcal{I})$. $\mathcal{I} \subseteq 2^X$ is the set of *independent subsets* of the universe X .¹ For our purposes, it will suffice to cover the semantic meaning of k -uniform matroids, which is codified as follows:

- (1) All subsets S of a feasible solution T must also be feasible solutions.
- (2) Every $T \subset X, |T| = k$ is a feasible solution and is maximal in the sense that no superset $T \subset R \subset X$ is feasible.

For general matroids, there exists a $1/2$ ratio for greedy maximization of submodular functions due to Fisher et al. [6]. This is a special case of their $1/(p+1)$ ratio for the intersection of p matroids.

Submodularity. The submodularity condition states that given any subsets $S \subset T$ of a universe X , the marginal gain of any $x \in X$ does not increase as the cardinality increases:

$$f(T \cup \{x\}) - f(T) \leq f(S \cup \{x\}) - f(S)$$

This formally encodes the idea of diminishing returns. Leskovec et al. exploited this property to show a data-dependent bound in terms of the marginal gain of the top- k un-selected elements [13], which was generalized to the adaptive case [8].

To the best of our knowledge, the only generally applicable relaxation of this constraint is the work of Wang et al. [26], who define a ratio in terms the *elemental curvature* of a function, which encodes the degree with which a function may break submodularity.

1.1.2 Alternate Problems & Algorithms. The $1 - 1/e$ ratio has shown surprising generality, with proofs that it holds for maximization of sequence functions [27] (and references) and adaptive stochastic maximization of functions that are submodular in expectation [8], among others. However, not all adjacent work relies on the same naïve greedy method. To obtain a bound on the relaxation of monotonicity, Buchbinder et al. [3] proposed a “double-greedy” algorithm with a $1/3$ (deterministic) or $1/2$ (randomized) ratio. For maximization on an intersection of $p \geq 2$ matroids, Lee et al. showed a $1/(p + \epsilon)$, $\epsilon > 0$ ratio for a local search method [12].

Vondrak et al. proposed a continuous greedy algorithm with a $(1/c)(1 - e^{-c})$ ratio for general matroids [25], where c is the total curvature of the function. An augmentation of this method has been shown to obtain a $(1 - c/e)$ -approximation for single matroids [23], along with an analogue for supermodular minimization. We remark that, while it exhibits a better ratio, this comes with a corresponding increase in complexity of the algorithm.

1.1.3 Curvature-Based Ratios. Conforti & Cornuéjols [5] introduced the idea of *total curvature* later used by Sviridenko et al. for their $(1 - c/e)$ ratio.

DEFINITION 1 (TOTAL CURVATURE). Given a monotone non-decreasing submodular function f defined on a matroid $\mathcal{M} = (\mathcal{I}, X)$, the total curvature of f is

$$c = \max_{j \in X} \left\{ 1 - \frac{f(X) - f(X \setminus \{j\})}{f(\{j\}) - f(\emptyset)} \right\}$$

Using this definition, they arrived at a $1/(1 + c)$ approximation for general matroids, which reduces to $\frac{1}{c}(1 - e^{-c})$ for maximization

¹For a complete treatment on matroids and associated theory, see Oxley [21].

on uniform matroids. Recently, Wang et al. [26] extended this idea by introducing the *elemental curvature* α of a function f :

DEFINITION 2 (ELEMENTAL CURVATURE). *The elemental curvature of a monotone non-decreasing function f is defined as*

$$\alpha = \max_{S \subseteq X, i, j \in X} \frac{f_i(S \cup \{j\})}{f_i(S)}$$

where $f_i(S) = f(S \cup \{i\}) - f(S)$.

While the resulting ratio (Theorem 1.1) is not as clean as that of prior work, this ratio is well-defined for non-submodular functions.

THEOREM 1.1 (WANG ET AL. [26]). *For a monotone non-decreasing function f defined on a k -uniform matroid \mathcal{M} , the greedy algorithm on \mathcal{M} maximizing f produces a solution satisfying*

$$\left[1 - \left(1 - A_k^{-1}\right)^k\right] f(S^*) \leq f(S)$$

where S is the greedy solution, S^* is the optimal solution, $A_k = \sum_{i=1}^{k-1} \alpha^i$ and α is the elemental curvature of f .

COROLLARY 1.2 (WANG ET AL. [26]). *When $\alpha = 1$, the ratio given by Theorem 1.1 converges to $1 - 1/e$ as $k \rightarrow \infty$.*

However, the ratios produced based on the elemental curvature rapidly converge to 0 for non-submodular functions. This behavior is shown in Figure 1. Even for $k = 25$, the ratio is effectively zero and therefore uninformative. In contrast, we show that our ratio produces significant bounds for two non-submodular functions, while still converging to the $1 - 1/e$ ratio for submodular functions.

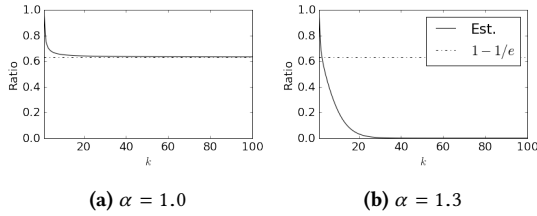


Figure 1: The ratio produced by Theorem 1.1 for (a) submodular and (b) non-submodular functions.

2 A RATIO FOR f NON-SUBMODULAR

In this section, we introduce a further extension to the notion of curvature: primal curvature. We derive a bound based on this, prove its equivalence to $1 - 1/e$ for submodular functions. Then, we extend the ratio to the adaptive case, which allows direct application to a number of problems modeled under incomplete knowledge. We adopt a problem definition similar to that of Wang et al. Specifically, our ratio applies to any problem that can be phrased as k -Uniform Matroid Maximization.

PROBLEM 1 (k -UNIFORM MATROID MAXIMIZATION). *Given a k -uniform matroid $\mathcal{M} = (X, \mathcal{I})$ and a monotone non-decreasing function $f : \mathcal{I} \rightarrow \mathbb{R}$, find*

$$S = \arg \max_{I \in \mathcal{I}} f(I)$$

2.1 Construction of the Ratio

As noted previously, the ratio given by elemental curvature rapidly converges to zero for non-submodular functions. We observe that this is due to the definition of α encoding the worst-case potential, and address this limitation by introducing the *primal curvature* of a function. Our definition separates the notion of rate-of-change from the global perspective imposed by elemental curvature.²

DEFINITION 3 (PRIMAL CURVATURE). *The primal curvature of a set function f is defined as*

$$\nabla_f(i, j | S) = \frac{f_i(S \cup \{j\})}{f_i(S)}$$

The global maximum primal curvature is equivalent to the elemental curvature of a function.

This shift from global to local perspective allows focus on the patterns present in real-world problem instances rather than limiting our attention to the worst-case scenarios.

A key observation of Wang et al's work is that the elemental curvature defines an upper bound on the change between $f(S)$ and $f(T)$, for some $S \subset T$, in terms of α and the marginal gain at S . The definition of primal curvature improves on this, giving an equivalence in terms of the *total primal curvature* Γ .

DEFINITION 4 (TOTAL PRIMAL CURVATURE). *The total primal curvature of $x \in X$ between two sets $A \subseteq B \subset X$ with $x \notin B$ is*

$$\Gamma(x | B, A) = \prod_{j=1}^r \nabla_f(x, b_j | S \cup \{b_1, b_2, \dots, b_{j-1}\})$$

where the b_j 's form an arbitrary ordering of $B \setminus A$ and $r = |B \setminus A|$.

We note that Γ can be interpreted as the *total change* in the marginal value of x from point A to point B . The following lemma illustrates this, as well as providing a useful identity.

LEMMA 2.1.

$$\Gamma(x | B, A) = \frac{f_x(A \cup B)}{f_x(A)}$$

PROOF. First, expand the product into its constituent terms:

$$\frac{f_x(A \cup \{b_1\})}{f_x(A)} \cdot \frac{f_x(A \cup \{b_1, b_2\})}{f_x(A \cup \{b_1\})} \dots \frac{f_x(A \cup B)}{f_x(A \cup \{b_1, b_2, \dots, b_{r-1}\})}$$

After cancelling, the statement immediately follows. \square

From this identity, we gain one further insight: the order in which elements are considered in Γ does not matter.

COROLLARY 2.2. *The product $\Gamma(x | B, A)$ is order-independent.*

PROOF. By Lemma 2.1, we see that each possible ordering reduces to the same fraction. Therefore, Γ is order-independent. \square

Using this, we can prove an equivalence between the change in total benefit and the sum of marginal gains taken with respect to S .

²The term *primal* is adopted primarily to distinguish this definition from prior work.

LEMMA 2.3. For a set function f and a pair of sets $S \subseteq T$,

$$f(T) - f(S) = \sum_{t=1}^r \Gamma(j_t \mid S_{t-1}, S) f_{j_t}(S)$$

where $r = |T \setminus S|$, $f_x(S) = f(S \cup \{x\}) - f(S)$ is the marginal gain, $S_t = S \cup \{j_1, j_2, \dots, j_t\}$, and $\Gamma(x \mid B, A)$ is the total primal curvature of f from A to B about x .

PROOF. Let j_1 be an arbitrary labeling of $T \setminus S$. Then we have:

$$f(T) - f(S) = f(S \cup \{j_1, j_2, \dots, j_r\}) - f(S) = \sum_{t=1}^r f_{j_t}(S_{t-1})$$

By the identity given in Lemma 2.1, we can write

$$f(T) - f(S) = \sum_{t=1}^r \Gamma(j_t \mid S_{t-1}, S) f_{j_t}(S)$$

Noting that $S \cup S_i = S_i$. Thus, the statement is proven. \square

With this lemma, we can now construct the ratio.

THEOREM 2.4. For a monotone non-decreasing function f defined on a k -uniform matroid $\mathcal{M} = (X, \mathcal{I})$, the greedy algorithm on \mathcal{M} maximizing f produces a solution satisfying

$$\left[1 + \left(\frac{f(S^+)}{f(S)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, S) + \epsilon_t) \right]^{-1} f(S^*) \leq f(S)$$

where S is the greedy solution, S^+ is the greedy solution for an identical problem if a $k+1$ -uniform supermatroid \mathcal{M}^+ of \mathcal{M} is well-defined, S^* is the optimal solution on \mathcal{M} , and $\hat{\Gamma}(i, S)$ is an estimator satisfying:

$$\forall T \in \mathcal{I}, S \subset T, i = |T \setminus S|, x \notin T \cup S : \Gamma(x \mid T, S) \leq \hat{\Gamma}(i, S) + \epsilon_i$$

PROOF. To begin, note that $f(S^*) \leq f(S^* \cup S)$ due to f monotone non-decreasing. Then, by Lemma 2.3 we have:

$$f(S^* \cup S) - f(S) = \sum_{t=1}^r \Gamma(x \mid S_{t-1}, S) f_{j_t}(S)$$

To make this computable with S^* unknown, we must make three relaxations. First, we remove the dependence on knowledge of S^* 's contents by substituting Γ with an estimator $\hat{\Gamma}$, satisfying the relation stated above.

$$f(S^*) - f(S) \leq f(S^* \cup S) - f(S) \leq \sum_{t=1}^r (\hat{\Gamma}(t-1, S) + \epsilon_{t-1}) f_{j_t}(S)$$

For the case of $t = 1$, we define $\hat{\Gamma}(0, S) = 1$ and $\epsilon_0 = 0$. Next, we remove the dependence on knowing $|S^* \setminus S|$ by relaxing r to k :

$$f(S^*) - f(S) \leq \sum_{t=1}^k (\hat{\Gamma}(t-1, S) + \epsilon_{t-1}) f_{j_t}(S) \quad (1)$$

Finally, we remove the dependence on knowing the elements j_t by noting that the $k+1$ st greedy selection g_{k+1} maximizes f_j at S :

$$f(S^*) - f(S) \leq f_{g_{k+1}}(S) \sum_{t=1}^k (\hat{\Gamma}(t-1, S) + \epsilon_{t-1}) \quad (2)$$

Then, rearranging terms we get

$$\left[1 + \left(\frac{f(S^+)}{f(S)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, S) + \epsilon_t) \right]^{-1} f(S^*) \leq f(S)$$

\square

When compared to traditional approximation ratios, this ratio has several obvious differences. First, it has dependencies on both the greedy solution and an extension of it to $k+1$ elements. This is both a strength and fundamental limitation of Theorem 2.4: it takes into account how much the greedy solution has converged toward negligible marginal gains, but also inhibits general analysis over all potential problem instances. Further, it requires that the supermatroid \mathcal{M}^+ be well-defined, though we remark that this is generally not a problem. In practice, most problems solved with greedy algorithms are k -element solutions on n -element spaces, with k typically much less than n .

2.2 Equivalence to the $1 - 1/e$ Ratio

We next show that under assumptions encoding the submodularity condition, the above is equivalent to the $1 - 1/e$ ratio as $k \rightarrow \infty$.

LEMMA 2.5. Given a sequence of ϵ_i -bounds $\hat{\Gamma}_i$ satisfying $\forall G : \hat{\Gamma}_i \geq \hat{\Gamma}(i, G)$, the greedy algorithm produces a k -element solution G_k satisfying

$$\left[1 - \left(1 - \frac{1}{\Lambda_k} \right)^k \right] f(S^*) \leq f(S)$$

where $\Lambda_k = \sum_{t=1}^k (\hat{\Gamma}_{t-1} + \epsilon_{t-1})$.

PROOF. We begin with Eqn. (2):

$$f(S^*) \leq f(S_l) + f_{g_{l+1}}(S) \sum_{t=1}^k (\hat{\Gamma}_{t-1} + \epsilon_{t-1})$$

for each $l \leq k$, where S_l denotes the l -element greedy solution. Substitute $\Lambda_k = \sum_{t=1}^k (\hat{\Gamma}_{t-1} + \epsilon_{t-1})$. Multiply both sides by $(1 - 1/\Lambda_k)^{k-l}$ and sum from $l = 1$ to $l = k$. The left-hand side becomes:

$$\Lambda_k \left[1 - \left(\frac{\Lambda_k - 1}{\Lambda_k} \right)^k \right] f(S^*) = \Lambda_k \left[1 - \left(1 - \frac{1}{\Lambda_k} \right)^k \right] f(S^*)$$

To obtain the right-hand side, separate $f(S_l) = \sum_{i=1}^l f_{g_i}(S_{i-1})$ into the marginal gain terms to produce the following in the body of the summation:

$$\left(\Lambda_k (1 - \Lambda_k^{-1})^{k-l} + \sum_{i=l+1}^k (1 - \Lambda_k^{-1})^{k-i} \right) f_{g_{l+1}}(S_l)$$

Summing this over l and employing the identity of the geometric series, this reduces to $\Lambda_k f(S_k) = \Lambda_k f(S)$ on the right-hand side. Thus, we obtain the relation

$$\left[1 - \left(1 - \frac{1}{\Lambda_k} \right)^k \right] f(S^*) \leq f(S)$$

\square

COROLLARY 2.6. For a submodular monotone non-decreasing function f , the following relation holds as $k \rightarrow \infty$:

$$(1 - 1/e) f(S^*) \leq f(S)$$

PROOF. For a submodular function, the primal curvature of any two elements u, v at any point T satisfies $\nabla(u, v \mid T) \leq 1$ by the definition of submodularity. Thus, with ϵ_i identically 0 we let $\hat{\Gamma}_i$ be 1. Then, the limit of $(1 - \frac{1}{\sum \hat{\Gamma}_i})^k = (1 - \frac{1}{k})^k$ as $k \rightarrow \infty$ is $1/e$. \square

Thus, we see that this ratio is a generalization of the classical $1 - 1/e$ approximation ratio that allows *specialization* of a ratio to the particular kind of problem instances being operated on. Further, the definition of total primal curvature illuminates why this ratio is capable of producing more useful bounds for non-submodular objectives than that of Wang et al: the Γ values encode a product of values that *may* converge to a limit, depending on problem instance, while the α bound uses $\prod_{t=0}^i \alpha = \alpha^i$ which does not converge for any $\alpha > 1$ (a condition which is implied by non-submodularity).

2.3 The Adaptive Ratio

We conclude this section by extending this ratio to the adaptive case where the decision made at each greedy step takes into account the outcomes of previous decisions. Briefly: in an adaptive algorithm, at each step the algorithm has a *partial realization* ψ consistent with the true realization Φ [8]. After each step, this partial realization is updated with the outcome of that step to form ψ' . The method for deciding the steps to take is termed a *policy*, with the greedy algorithm encoded as the greedy policy.

This representation supports the study of algorithms that operate with incomplete information and gradual revelation of the data. The initial motivation was described in terms of placement of sensors that may fail, and this technique has seen further use in studying networks with incomplete topology [15, 22], active learning under noise [9], and distributed representative subset mining [17].

We generalize our ratio to this case by defining the *adaptive* primal curvature of a function in terms of the partial realizations.

DEFINITION 5 (ADAPTIVE PRIMAL CURVATURE). *The primal curvature of an adaptive monotone non-decreasing function f is*

$$\nabla_f(i, j \mid \psi) = \mathbb{E} \left[\frac{\Delta(i \mid \psi \cup s)}{\Delta(i \mid \psi)} \mid s \in S(j) \right]$$

where $S(j)$ is the set of possible states of j and Δ is the conditional expected marginal gain [8].

This definition leads to the following theorem by similar arguments as Thm. 2.4.³

THEOREM 2.7. *For an adaptive monotone non-decreasing function f , the solution given by a greedy policy π selecting k elements satisfies:*

$$\left[1 + \left(\frac{f_{\text{avg}}(\pi^+)}{f_{\text{avg}}(\pi)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, \pi) + \epsilon_t) \right]^{-1} f_{\text{avg}}(\pi^*) \leq f_{\text{avg}}(\pi)$$

where π^+ is the greedy policy selecting $k + 1$ elements, π^* is the optimal k -element policy, and $\hat{\Gamma}(i, S)$ is an estimator satisfying

$$\Gamma(x \mid \psi', \psi) \leq \hat{\Gamma}(i, \pi) + \epsilon_i$$

$$\forall \psi \in \pi(\emptyset), \psi \subset \psi', i = |\text{dom}(\psi') \setminus \text{dom}(\psi)|, x \notin \text{dom}(\psi') \cup \text{dom}(\psi)$$

where $\text{dom}(\psi)$ is the set of elements observed in partial realization ψ .

³Proof in Appendix B.

3 ESTIMATING TOTAL PRIMAL CURVATURE

Having established our ratio, we now turn to the problem of constructing the estimator $\hat{\Gamma}$. Recall the relation it must satisfy:

$$\forall T \in \mathcal{I}, S \subset T, i = |T \setminus S|, x \notin T \cup S : \Gamma(x \mid T, S) \leq \hat{\Gamma}(i, S) + \epsilon_i$$

In plain English, this estimator acts as an upper bound on the total primal curvature values Γ with error margin ϵ_i , indexed by the size of the difference between T and S . Unfortunately, finding a deterministic, strict estimator for Γ does not seem feasible without also sacrificing the benefits gained by this ratio over prior work. That is, if we were to construct such a Γ that were a strict upper bound over all sequences, the result would align with the absolute worst case rather than with the bulk of possible cases. In such a scenario, it is unlikely that there would be any advantage to using the Γ over the α ratio of Wang et al. Therefore, we turn to a probabilistic estimator that captures the variance in Γ to provide a tunable guarantee based on patterns present in real-world data. In this section, we describe a solution based on Monte-Carlo simulation, which we use to construct an estimator with error margin $\epsilon_i = \epsilon$ that holds with probability δ . Algorithm 2 gives the description for our method.

Algorithm 1 $\tilde{\Gamma}$ Sequence Sampler

Input: Uniform Matroid $\mathcal{M} = (X, \mathcal{I})$, Greedy solution S, k

Output: Curvature sequence $Q = [\tilde{\Gamma}_i]$

```

1: Randomly select  $x \in X \setminus S; T \leftarrow S$ 
2:  $\gamma \leftarrow 1, Q \leftarrow [\gamma]$ 
3: for  $i = 1 \dots k$  do
4:   Randomly select  $y \in X \setminus T \setminus \{x\}$ .
5:    $T \leftarrow T \cup \{y\}$ 
6:    $\gamma \leftarrow \gamma \nabla_f(x \mid T, S)$ 
7:   Append  $\gamma$  to  $Q$ 
8: end for
9: return  $Q$ 
```

This algorithm begins by sampling a set of $\tilde{\Gamma}$ sequences, each of

Algorithm 2 Matching- ϵ Γ -Estimator

Input: Uniform Matroid $\mathcal{M} = (X, \mathcal{I})$, Greedy solution S, k , probability δ , relative stopping conditions α^*, β^* .

Output: Total Primal Curvatures $\hat{\Gamma}$, error margin ϵ

```

1:  $\alpha \leftarrow \infty, \beta \leftarrow \infty, s \leftarrow 10,000, A \leftarrow \emptyset, \epsilon_p \leftarrow \infty, \Gamma_p \leftarrow \infty$ 
2: repeat
3:    $A \leftarrow A \cup \{\text{Sample } s \text{ sequences with Alg. 1}\}$ 
4:    $s \leftarrow |A|$ 
5:    $\bar{\Gamma} \leftarrow \text{mean}(A)$ 
6:   Select  $\epsilon$  such that  $\Pr[Z \leq \bar{\Gamma} + \epsilon] \geq \delta$ ,  $Z$  a multivariate normal random variable with mean  $\bar{\Gamma}$  and covariance  $\text{cov}(A)$ .
7:    $\alpha \leftarrow |(\epsilon - \epsilon_p)/\epsilon|, \beta \leftarrow \|\bar{\Gamma} - \Gamma_p\|_2 / \|\bar{\Gamma}\|_2$ 
8:    $\epsilon_p \leftarrow \epsilon, \Gamma_p \leftarrow \bar{\Gamma}$ 
9: until  $\alpha \leq \alpha^*$  and  $\beta \leq \beta^*$ 
10: return  $\bar{\Gamma}, \epsilon$ 
```

which may be constructed in parallel by Algorithm 1. These values are used to estimate the mean sequence $\bar{\Gamma}$. As these sequences are independent and identically distributed, by the central limit theorem, we know that this converges to a (multivariate) normal distribution. We construct the distribution using the mean and

covariance of the sampled sequences, and use the method of Genz & Bretz [7] to estimate ϵ . This process is repeated until both the error margin and mean have converged. While we do not have a proof of the rate of convergence, in our experiments we have observed that it typically converges after two or three iterations.

3.1 Worst-Case Biased Sampling

The sampling method used in Algorithm 1 encodes the assumption that the difference $S^* \setminus S$ is equally likely to be any set. However, it is reasonable to conclude that this is not the case. To illustrate this, we introduce an alternate definition of the (in)dependence of elements $x, y \in X$. Note that this is *not* the same as the (in)dependence of sets on the matroid \mathcal{M} . Using this, we describe the worst-case behaviors for sub- and super-modular functions.

DEFINITION 6. Two elements $x, y \in X$ are dependent on a matroid $\mathcal{M} = (X, \mathcal{I})$ under a function f iff $\exists T \in \mathcal{I}, x, y \notin T : \nabla_f(x, y|T) \neq 1$. The set of elements dependent on an element x is written $D(x)$.

Submodular. The worst case ratio for a fixed solution S occurs when the sequence of $\hat{\Gamma}$ is maximal. Note that the maximal Γ for a submodular function is 1, which occurs when each element $y \in T \setminus S$ is independent of the root element x . The worst case for submodular functions occurs when every possible sequence is independent of the root. Further, this case is expected to be similar to the optimal, as selection of dependent elements will only cause the marginal gain to decrease. Thus, we expect that *for submodular functions, the optimal is biased towards independent elements*.

Non-Submodular. For a non-submodular function, there is no upper or lower limit on the value of ∇_f , and thus no limit on Γ . The maximal sequence consists of each dependent element with $\nabla_f \geq 1$, followed by independent elements. In the special case of supermodular functions, the worst case is a sequence of dependent elements. What's more: selection of dependent elements with a supermodular weight increases the weight of the selected elements. Therefore, we expect that *for supermodular functions, the optimal is biased towards dependent elements*. No such claim can be made for simultaneously non-submodular and non-supermodular functions.

Moreover, we remark that many datasets are heavily biased towards highly-independent sequences. To see this, suppose that $\Pr[A \in D(B)] = p$ for random elements A, B . It is clear that even for relatively large p , the probability of randomly sampling a worst-case sequence (i.e. every element dependent on another element) for a supermodular function rapidly approaches zero as k increases, while for a submodular function it approaches 1. As a result, we observe that is necessary to bias the sampling of sequences in order to better represent the worst-case scenario for varying problems.

We therefore introduce a bias parameter b fixing the probability of dependence for the sampling procedure. $b = 0$ then always produces fully independent sequences corresponding to the worst case for submodular functions, and $b = 1$ produces fully dependent sequences corresponding to the worst case for supermodular functions. However, for functions that are neither supermodular nor submodular, there is no apparent choice of b . Therefore, we estimate the ratio for a range of values on $[0, 1]$.

3.2 The r -Gap Ratio

We conclude this section by introducing a corollary that will find use in our subsequent analysis of the ratio. Recall that in Eqn. 1, our knowledge of $r = |S^* \setminus S|$ is relaxed from complete (knowing r exactly) to none. Suppose instead that the distribution of possible r -values ρ is known. Then, we have the following ratio:

COROLLARY 3.1. For a monotone non-decreasing function f defined on a k -uniform matroid $\mathcal{M} = (X, \mathcal{I})$, the greedy algorithm on \mathcal{M} maximizing f produces a solution satisfying

$$\left[1 + \left(\frac{f(S^*)}{f(S)} - 1 \right) \mathbb{E} \left[\sum_{t=0}^{r-1} (\hat{\Gamma}(t, S) + \epsilon_t) \right] \right]^{-1} f(S^*) \leq f(S)$$

where S is the greedy solution, S^+ is the greedy solution for an identical problem if a $k+1$ -uniform supermatroid \mathcal{M}^+ of \mathcal{M} is well-defined, S^* is the optimal solution on \mathcal{M} , $\hat{\Gamma}(i, S)$ is an estimator satisfying the following relation:

$$\forall T \in \mathcal{I}, S \subset T, |T \setminus S| = i, x \notin T \cup S : \Gamma(x | T, S) \leq \hat{\Gamma}(i, S) + \epsilon_i$$

and the expectation is taken w.r.t. a distribution ρ of values of r .

While at this time we do not have a method of estimating ρ without knowing a set of optimal solutions, this ratio will be valuable for explaining the source of the gap between the Γ (and α) ratio and real-world performance in the next section.

4 EVALUATION

In this section, we evaluate the Γ ratio on three problems. First, to the submodular problem of Influence Maximization on the IC and LT models. This establishes a baseline, allowing us to examine the behavior of the ratio within the context of an algorithm with known worst and expected case behavior. The worst-case ratio is shown to converge to $1 - 1/e - \epsilon$, and we explore the factors that lead to this behavior. Further, if we consider a possible distribution of r based on the observed differences between greedy and exact solutions, we show that this ratio can be improved by nearly 20%.

We next examine two non-submodular problems. First, Robust Influence Maximization: a nonlinear, non-submodular combination of Influence Maximization sub-problems. We show that although the $\hat{\Gamma}$ values are very similar to those of SSA for Influence Maximization, the ϵ values reduce the worst-case ratio to $1 - 1/\sqrt{e}$. Further, this problem is shown to behave like a submodular function with respect to sampling bias. Finally, we examine the Max-Crawling problem, which demonstrates an application of our adaptive ratio. We find a ratio of nearly 1/2 for this problem under a realistic, non-submodular model of user behavior that is consistent across datasets, and that it exhibits pseudo-supermodular bias behavior.

The ratio estimator is implemented in Rust and is available along with all experimental data and supplemental code⁴. For submodular objective functions, we additionally placed an upper bound on $\hat{\Gamma}(i, G) + \epsilon$ to prevent it from exceeding 1 due to variance. All evaluations are run with $\delta = 0.95$, $\alpha = 0.05$, $\beta = 0.01$. We estimate the ratio for each $b = [0, 0.1, \dots, 1]$ and report the worst-case ratio predicted by our method, along with a ratio assuming that all settings of b and r are equally likely, which encodes a scenario

⁴<https://github.com/emallson/ava.git>

	Dataset	True	Worst Case	Uniform
Kempe	Gr-Qc	0.99	0.625	0.780
	NetPHY	1.0	0.628	0.770
IMM	Gr-Qc	1.0	0.665	0.781
	NetPHY	1.0	0.628	0.770
SSA	Gr-Qc	0.91	0.627	0.806
	NetPHY	0.94	0.611	0.781
Robust IM	Gr-Qc	-	0.508	0.704
	NetPHY	-	0.390	0.625
Max-Crawling	Gr-Qc	-	0.498	0.700
	NetPHY	-	0.478	0.682

Table 1: True, worst case, and uniform expectation ratios for each problem. To conserve space, we list the worst case ratios for IMM and SSA over both IC and LT models. The uniform expectation ratio is computed assuming a uniform distribution over bias and r .

	Nodes	Edges
Gr-Qc [14]	5242	14496
NetPHY [4]	37149	180826

Table 2: Datasets used in experiments.

in which we have no knowledge of the optimal solution. For Influence Maximization, we additionally report exact ratios. Table 1 summarizes our results and Table 2 details the datasets we test on.

4.1 Baseline: Influence Maximization

We begin our evaluation by examining the Influence Maximization problem, which asks us to find a set of users S maximizing the expected number of users activated on a (often: social) network.

PROBLEM 2 (INFLUENCE MAXIMIZATION). *Given a graph $G = (V, E)$ and an influence function $\sigma : 2^V \rightarrow \mathbb{R}$ mapping seed sets to the expected number of activated nodes, find*

$$S = \arg \max_{T \subset V, |T| \leq k} \sigma(T)$$

This problem was originally solved directly with a greedy algorithm using Monte Carlo sampling by Kempe et al. [11]. Recently, an alternative phrasing in terms of solving the Maximum Coverage problem on *Reverse Influence Samples* of the network has come to prominence as a more efficient solution [1]. We estimate the approximation quality of three algorithms for the Influence Maximization problem: Kempe et al. [11], IMM [24], and SSA [19]. We use a $(1 - \epsilon)$ -approximate solution [16] to establish the true ratio for comparison. Both IMM and SSA are $(1 - 1/e - \epsilon)$ -approximate with a probability $1 - \delta$, where ϵ encodes a user-provided sampling error. In all cases, ϵ is fixed at 0.02 and $\delta = 1/n$, where n is the number of nodes in the graph. We use the standard edge weighting scheme of $w_{uv} = 1/d_{in}(v)$, where $d_{in}(v)$ is the in-degree of v .

Figures 2-3 show the r -Gap ratio for each method. We find that the ϵ values are sufficiently large to cause the \hat{r} terms to reach the threshold of 1. However, the worst-case ratio remains above the $1 - 1/e - \epsilon \approx 0.61$ lower bound in all cases. The plots for the Gr-Qc dataset are substantially similar, and are therefore omitted. Further, we note that while these plots are similar for each implementation, the observed \hat{r} sequences shown in Figure 4 differ for each: IMM has higher estimates of \hat{r} , while after ϵ is included they

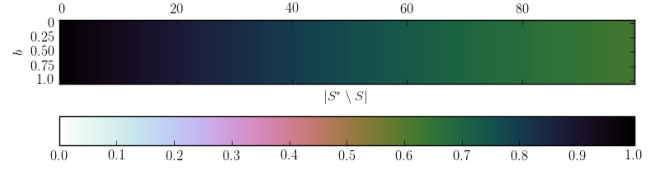
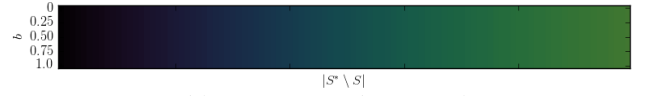


Figure 2: The r -Gap Ratio for Kempe on the Gr-Qc dataset under the IC model. (Min: 0.625)



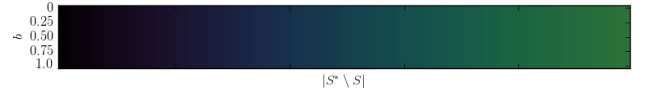
(a) IMM, IC model (Min: 0.611)



(b) SSA, IC model (Min: 0.628)



(c) IMM, LT model (Min: 0.649)



(d) SSA, LT model (Min: 0.661)

Figure 3: The r -Gap Ratio on NetPHY.

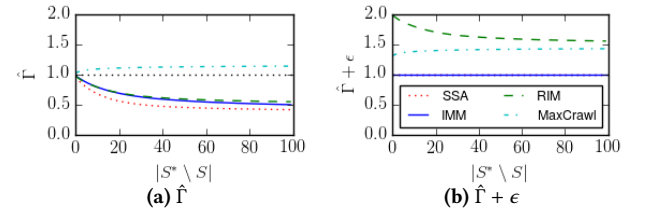


Figure 4: Worst-case \hat{r} sequence estimates for each problem. The line at 1.0 corresponds to the conditions of Corollary 2.6.

perform near-identically and both converge to a sequence of 1's. Figure 5 shows the covariance matrix for each \hat{r} . Note that the maximal variance is nearly 10x as large as the minimal. However, the computational limitations on estimation of \hat{r} prevent us from exploiting this to construct tighter bounds: running [7] takes anywhere from 180ms to several seconds depending on dimensionality, which quickly becomes impractical given the number of function calls convergence required in our testing.

Based on the $1 - \epsilon$ solutions produced by TipTop, we note that the gap never exceeds 60 elements. In this case, we can construct a rudimentary estimate of ρ , which gives an r -Gap ratio of at worst 0.723, with a ratio of 0.842 if r is assumed uniform on $[0, 60]$ – both on IMM under the IC model. This is a gain of 18% over the general worst case and of 9% of the general uniform case.

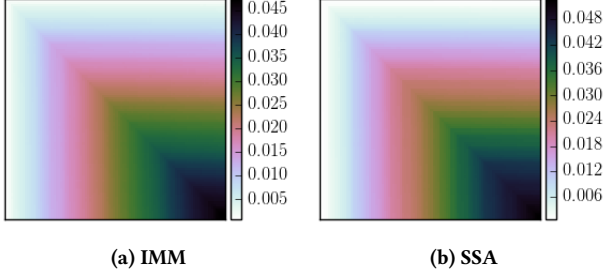
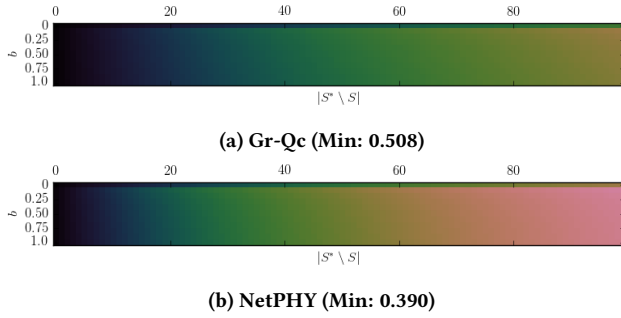


Figure 5: Worst-case covariance estimates for IMM and SSA.

4.2 The $1 - 1/\sqrt{e}$ -ratio for Robust Influence Maximization

Figure 6: The r -Gap Ratio for the Robust Influence Maximization problem over the IC and LT models with $\zeta = 10^{-16}$.

We now turn our attention to the first non-submodular problem we will consider: Robust Influence Maximization. This problem extends Prob.2 to a set of potential influence functions, and asks us to find a solution that maximizes the minimum fraction of influence gained. This guards against the worst-case scenario: maximizing for a model σ at the expense of σ' , only to ultimately find σ' to be the true model [10].

PROBLEM 3 (ROBUST INFLUENCE MAXIMIZATION). *Given a set Σ of influence functions, maximize the objective*

$$\rho(S) = \min_{\sigma \in \Sigma} \frac{\sigma(S)}{\sigma(S_\sigma^*)}$$

subject to a cardinality constraint $|S| \leq k$, where S_σ^ is an optimal solution for model σ .*

The min term prevents submodularity from holding in general, even if all $\sigma \in \Sigma$ are submodular. However, it is monotone non-decreasing and the problem forms a uniform matroid, similarly to Prob. 2. In our evaluation, we estimate $\sigma(\cdot)$ with SSA, using its ratio to estimate the value of the optimal.

Note that it is possible for ∇f to be ill-defined on this problem: if for a node u , $\rho(S \cup \{u\}) - \rho(S) = 0$, then u has no marginal gain under the current minimum model σ . However, if adding another v causes this model to change to σ' , then it is possible that $\rho(S \cup \{u, v\}) - \rho(S \cup \{v\}) \neq 0$. To address this, we introduce a small term ζ into the definition of primal curvature:

$$\nabla f(u, v | S) = \frac{f_u(S \cup \{v\}) + \zeta}{f_u(S) + \zeta} \approx \frac{f_u(S \cup \{v\})}{f_u(S)}$$

Using this definition, we get that $\Gamma(x | B, A) = (f_u(B \cup A) + \zeta)/(f_u(A) + \zeta)$ by the same proof as Lemma 2.1, and a ratio of

$$\left[1 + \left(\frac{f(S^+) + \zeta}{f(S)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, S) + \epsilon_t) - k\zeta \right]^{-1} f(S^*) \leq f(S)$$

after propagating the ζ term through the proof of Theorem 2.4.

Figure 6 shows this ratio. The worst case ratio is $0.39 \approx 1 - 1/\sqrt{e}$ across all bias and r values. Further, we can see in each heatmap that greedy Robust Influence Maximization exhibits pseudo-submodular worst-case behavior with respect to the bias parameter (that is, $b \approx 0$ gives the worst-case ratio). Examining Figure 4 we see that it has a similar $\hat{\Gamma}$ sequence to SSA. Though this is not surprising given the implementation details, these two facts highlight a strength of our ratio: for scenarios where the worst case can be arbitrarily bad, it describes what we can reasonably expect the behavior to be. In this case, we can reasonably expect using SSA sampling to solve RIM with submodular influence models will exhibit roughly submodular behavior.

While useful, this does come with caveats. Most notably, since this method specializes the ratio to the given data, it may not generalize across datasets. This limitation is exemplified by the difference in ratios for the NetPHY and Gr-Qc datasets. However, we remark that this difference can also be explained by the different in scales of the graph: selecting k elements on Gr-Qc covers a much smaller fraction than on NetPHY, so we may reasonably expect that the $f(S^+)/f(S)$ term will be larger. Generalized claims are further inhibited by the difficulty of describing the interactions of the $\hat{\Gamma}$ sequence and f -ratio, making the general behavior of the ratio difficult to predict as a function of k .

4.3 The $1/2$ -ratio for Max-Crawling

The final problem we consider is a maximization variant of the Min-Friending problem considered by Li et al. [15], which describes the task of modeling an optimal “socialbot” conducting reconnaissance on a social network. The bot is capable of making friends on the network, and seeks to do so in a sequence which allows it to optimally infiltrate a targeted set of nodes.

PROBLEM 4 (ADAPTIVE MAX-CRAWLING). *Given a partially-known social network $G = (V, E)$, a target set T , an acceptance function $\alpha(u | \psi)$, and a budget $k \in \mathbb{Z}^+$, find an adaptive strategy π such that after k friend requests the following objective function is maximized:*

$$f(\pi) = \mathbb{E} \left[\sum_{u \in F} B_F(u) + \sum_{v \in F'} B_{F'}(v) + \sum_{(u, v) \in R} B_R(u, v) \right]$$

where F is the set of friends gained by the socialbot, F' is the set of friends of $u \in F$, R is a set of edges whose state (existing or non-existing) is revealed over the course of the attack, and B_ are monotone non-decreasing benefit functions for each.*

This problem is *adaptive*: each step taken depends on the outcome of the previous step (e.g. whether the friend request was accepted, and what edges were revealed). Li et al. proved that when network topology is not a factor in acceptance of friend requests, f_{avg} is adaptive submodular. However, they note this condition does not match observations well, and give an acceptance function

fitted to data drawn from socialbot experiments on Facebook [2], termed the *Expected Triadic Closure (ETC)* model:

$$\alpha(u \mid \psi) = \rho_1 \log(1 + \mathbb{E}[|N(u) \cap N(s)|]) + \rho_0$$

where s is the socialbot, $N(x)$ is the set of neighbors of x , and $\rho_1 = 0.22806$, $\rho_0 = 0.18015$ are constants. While the greedy algorithm was observed to perform well under this model – even relative to the fixed-acceptance model – no approximation ratio was given.

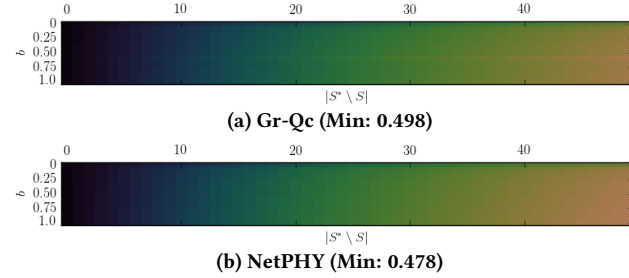


Figure 7: The r -Gap Ratio for Max-Crawling under the ETC model.

Figure 7 shows the ratio obtained for this model. Notably, the ratio is very close to $1/2$ even in the worst case, and well above it in the expected case. In contrast to the Robust Influence Maximization problem, Max-Crawling exhibits supermodular behavior with respect to the bias. This shows the importance of considering the bias, as the size of the dependent sets is small (on average about 100 nodes) relative to the overall size of the networks – without biasing the sampling, the worst-case ratio would be off by 10%.

5 CONCLUSION & FUTURE WORK

In this paper, we presented a method for estimating the approximation ratio of greedy maximization that works transparently for both submodular and non-submodular functions, in addition to a variant supporting adaptive greedy algorithms. This ratio reduces to at worst $1 - 1/e$ as $k \rightarrow \infty$ for submodular functions, and is shown to provide performance bounds for non-submodular maximization. Using our method, we show that greedily solving the Robust Influence Maximization problem is at least $1 - 1/\sqrt{e}$ -approximate on common datasets, while greedily solving the Max-Crawling problem with a real-world acceptance model is $1/2$ -approximate.

While we have demonstrated the utility of our technique for understanding the performance of non-submodular maximization, there remains room for further development. Relaxations of the uniformity and monotonicity conditions have found widespread use for submodular functions, and we expect that relaxing them for this ratio would likewise be generally useful. Further, there are two primary avenues by which we foresee improvement of the \hat{r} estimator: better methods for sampling relevant \vec{I} sequences, and better ways of estimating the error margin \vec{e} . These two approaches are closely related, and we expect that future work addressing these limitations will not only improve results for non-submodular maximization, but will place this method as a valuable tool for analysis of submodular maximization as well.

REFERENCES

- [1] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. Maximizing Social Influence in Nearly Optimal Time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* (2014) (SODA '14). SIAM.
- [2] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. The Socialbot Network: When Bots Socialize for Fame and Money. In *Proceedings of the 27th Annual Computer Security Applications Conference* (2011) (ACSAC '11). ACM, 93–102.
- [3] N. Buchbinder, M. Feldman, J. Naor, and R. Schwartz. A Tight Linear Time $(1/2)$ -Approximation for Unconstrained Submodular Maximization. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science (FOCS)* (2012-10).
- [4] W. Chen, Y. Wang, and S. Yang. 2009. Efficient influence maximization in social networks. In *KDD '09*. ACM, New York, NY, USA, 199–208.
- [5] Michele Conforti and Gérard Cornuéjols. 1984. Submodular Set Functions, Matroids and the Greedy Algorithm: Tight Worst-Case Bounds and Some Generalizations of the Rado-Edmonds Theorem. *Discrete Applied Mathematics* 7 (1984).
- [6] Marshall L. Fisher, George L. Nemhauser, and Laurence A. Wolsey. An Analysis of Approximations for Maximizing Submodular Set functions II. In *Polyhedral Combinatorics*. Springer, 73–87.
- [7] Alan Genz and Frank Bretz. 2009. *Computation of multivariate normal and t probabilities*. Vol. 195. Springer Science & Business Media.
- [8] Daniel Golovin and Andreas Krause. 2011. Adaptive Submodularity: Theory and Applications in Active Learning and Stochastic Optimization. *Journal of Artificial Intelligence Research* 42 (2011), 427–486.
- [9] Daniel Golovin, Andreas Krause, and Debajyoti Ray. Near-Optimal Bayesian Active Learning with Noisy Observations. In *Advances in Neural Information Processing Systems* 23, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Eds.). Curran Associates, Inc., 766–774.
- [10] Xinran He and David Kempe. 2016. Robust Influence Maximization. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, New York, NY, USA, 885–894.
- [11] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the Spread of Influence Through a Social Network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2003) (KDD '03). ACM.
- [12] Jon Lee, Maxim Sviridenko, and Jan Vondrák. Submodular Maximization over Multiple Matroids via Generalized Exchange Properties. 35, 4 (2007), 795–806.
- [13] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. Cost-Effective Outbreak Detection in Networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2007) (KDD '07). ACM, 420–429.
- [14] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>. (June 2014).
- [15] Xiang Li, J. David Smith, Thang N. Dinh, and My T. Thai. 2016. Privacy Issues in Light of Reconnaissance Attacks with Incomplete Information. In *Proceedings of the 2016 IEEE/WIC/ACM International Conference on Web Intelligence*. IEEE/WIC/ACM.
- [16] Xiang Li, J. David Smith, Thang N. Dinh, and My T. Thai. 2017. Why approximate when you can get the exact? Optimal Targeted Viral Marketing at Scale. In *IEEE INFOCOM 2017*. IEEE.
- [17] Baharan Mirzasoleiman, Amin Karbasi, Rik Sarkar, and Andreas Krause. Distributed Submodular Maximization: Identifying Representative Elements in Massive Data. In *Advances in Neural Information Processing Systems* 26, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 2049–2057.
- [18] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. 1978. An Analysis of Approximations for Maximizing Submodular Set functions – I. *Mathematical Programming* 14, 1 (1978), 265–294.
- [19] Hung T. Nguyen, Thang N. Dinh, and My T. Thai. Stop-and-Stare: Optimal Sampling Algorithms for Viral Marketing in Billion-Scale Networks. In *Proceedings of the 2016 ACM SIGMOD International Conference on Management of Data* (2016) (SIGMOD '16).
- [20] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer.
- [21] James G. Oxley. 1992. *Matroid theory*. Oxford University Press, New York.
- [22] L. Seeman and Y. Singer. Adaptive Seeding in Social Networks. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science* (2013-10). 459–468.
- [23] Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal Approximation for Submodular and Supermodular Optimization with Bounded Curvature. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms* (2015) (SODA '15). Society for Industrial and Applied Mathematics, 1134–1148.
- [24] Youze Tang, Yanchen Shi, and Xiaokui Xiao. Influence Maximization in Near-Linear Time: A Martingale Approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data* (2015) (SIGMOD '15). ACM.
- [25] Jan Vondrák. 2010. Submodularity and curvature: the optimal algorithm. *RIMS Kokyuroku Bessatsu B* 23 (2010), 253–266.
- [26] Zengfu Wang, Bill Moran, Xuezhi Wang, and Quan Pan. 2014. Approximation for maximizing monotone non-decreasing set functions with a greedy method. *Journal of Combinatorial Optimization* 31, 1 (Jan. 2014), 29–43.
- [27] Z. Zhang, E. K. P. Chong, A. Pezeshki, and W. Moran. 2016. String Submodular Functions With Curvature Constraints. 61, 3 (2016), 601–616.

A NON-MATCHING ϵ_i

A further extension to Algorithm 2 is a method to compute a vector $\vec{\epsilon} = [\epsilon_1 \dots \epsilon_k]$ of error margins. The justification for this is straightforward: unless all variables have equal variances, which is exceedingly unlikely, then the ratio will be tighter by allowing those with lower variance to have a correspondingly lower ϵ_i . This leads us to the following convex optimization problem:

$$\min \|\vec{\epsilon}\|_2 \quad (3)$$

$$s.t. \Pr \left[-\infty \leq X \leq \vec{\Gamma} + \vec{\epsilon} \right] \geq \delta \quad (4)$$

which can be solved by a number of techniques by employing the method of (Augmented) Lagrangian Multipliers (see e.g. [20] for details) – provided that each gradient can be computed. However, Eq. (4) is difficult to compute. The exact form of this CDF is [7]:

$$\Phi_k(\mathbf{a}, \mathbf{b}; \Sigma) = \frac{1}{\sqrt{|\Sigma|(2\pi)^k}} \int_{a_1}^{b_1} \dots \int_{a_k}^{b_k} e^{-\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}} d\mathbf{x}_k \dots d\mathbf{x}_1 \quad (5)$$

with $\mathbf{a} = [-\infty, \dots, -\infty]$ $\mathbf{b} = \vec{\epsilon}$, and Σ the covariance matrix. While there is no evaluable closed-form of (5), there is significant literature on numerical computation (see [7] and references) which is used both with matching and non-matching ϵ_i . Further, if we let \mathbf{a} and Σ be constant, the partial derivative $\frac{\partial \Phi_k}{\partial x_i}$ is equivalent to solving the subproblem with the variable x_i fixed at b_i .

LEMMA A.1.

$$\frac{\partial \Phi_k}{\partial b_i} = \frac{1}{\sqrt{|\Sigma|(2\pi)^k}} \int_{a_1}^{b_1} \dots \int_{a_i}^{b_i} e^{-\frac{1}{2} \mathbf{y}_i^T \Sigma^{-1} \mathbf{y}_i} d\mathbf{x}_i$$

where $\mathbf{a}_i, \mathbf{b}_i, \mathbf{x}_i$ are their respective vectors with the i 'th element omitted and \mathbf{y}_i is equal to \mathbf{x} with x_i fixed to b_i .

PROOF. To begin, note that the partial derivative w.r.t. b_i is constant w.r.t. the first $i-1$ integrals. Therefore, without loss of generality we focus on the sub-equation

$$\frac{\partial}{\partial b_i} \int_{a_i}^{b_i} \dots \int_{a_k}^{b_k} g(\mathbf{x}, \Sigma) d\mathbf{x}_k \dots d\mathbf{x}_i$$

where $g(\mathbf{x}, \Sigma) = \exp(-1/2 \mathbf{x}^T \Sigma^{-1} \mathbf{x})$. Write

$$h(x_i) = \int_{a_{i+1}}^{b_{i+1}} \dots \int_{a_k}^{b_k} g(\mathbf{x}, \Sigma) d\mathbf{x}_k \dots d\mathbf{x}_{i+1}$$

Then, we have the following by the fundamental theorem of calculus

$$\frac{\partial}{\partial b_i} \int_{a_i}^{b_i} h(x_i) dx_i = h(b_i)$$

Substituting this back into Φ_k we get

$$\frac{\partial \Phi_k}{\partial b_i} = \frac{1}{\sqrt{|\Sigma|(2\pi)^k}} \int_{a_1}^{b_1} \dots \int_{a_i}^{b_i} g(\mathbf{y}_i, \Sigma) d\mathbf{x}_k \dots d\mathbf{x}_{i+1} d\mathbf{x}_{i-1} \dots d\mathbf{x}_1$$

where $\mathbf{y}_i = [x_1, \dots, b_i, \dots, x_k]$, which is equivalent to the statement of the lemma. \square

Therefore, we are able to compute the gradient $\nabla \Phi_k$ in much the same way as we compute Φ_k (that is, numerically via [7]).

B PROOF OF THE ADAPTIVE RATIO

While in [8], f_{avg} is defined with respect to policies, we re-define it with respect to partial realizations for convenience of notation:

$$f_{\text{avg}}(\psi) = \mathbb{E}[f(\text{dom}(\psi), \Phi) \mid \Phi \sim \psi]$$

DEFINITION 7 (ADAPTIVE PRIMAL CURVATURE). *The primal curvature of an adaptive monotone non-decreasing function f is*

$$\nabla_f(i, j \mid \psi) = \mathbb{E} \left[\frac{\Delta(i \mid \psi \cup s)}{\Delta(i \mid \psi)} \mid s \in S(j) \right]$$

where $S(j)$ is the set of possible states of j and Δ is the conditional expected marginal gain [8].

For notational clarity, we also define the fixed adaptive primal curvature in terms of a single state $s \in S(j)$:

$$\nabla'(i, s \mid \psi) = \frac{\Delta(i \mid \psi \cup s)}{\Delta(i \mid \psi)}$$

DEFINITION 8 (ADAPTIVE T.P.C.). *Let $\psi \subset \psi'$ and $\psi \rightarrow \psi'$ represent the set of possible state sequences leading from ψ to ψ' . Then the adaptive total primal curvature is*

$$\Gamma(i \mid \psi', \psi) = \mathbb{E} \left[\prod_{s_j \in Q} \nabla'(i, s_j \mid \psi \cup \{s_1, \dots, s_{j-1}\}) \mid Q \in \psi \rightarrow \psi' \right]$$

LEMMA B.1.

$$\Gamma(i \mid \psi', \psi) = \frac{\Delta(i \mid \psi')}{\Delta(i \mid \psi)}$$

PROOF. Fix a sequence $Q \in \psi \rightarrow \psi'$ of length r . Then, expanding the product we obtain

$$\frac{\Delta(i \mid \psi \cup \{s_1\})}{\Delta(i \mid \psi)} \cdot \frac{\Delta(i \mid \psi \cup \{s_1, s_2\})}{\Delta(i \mid \psi \cup \{s_1\})} \dots \frac{\Delta(i \mid \psi')}{\Delta(i \mid \psi' \setminus \{s_{r-1}\})}$$

If we take the expectation of this w.r.t. the possible sequences Q , we obtain the same ratio regardless of Q , and therefore the claim holds trivially. \square

LEMMA B.2. *For an adaptive monotone non-decreasing function f and a pair of realizations $\psi \subset \psi'$*

$$f_{\text{avg}}(\psi') - f_{\text{avg}}(\psi) = \mathbb{E} \left[\sum_{t=1}^r \Gamma(j_t \mid \psi_{t-1}, \psi) \Delta(j_t \mid \psi) \mid Q \in \psi \rightarrow \psi' \right]$$

where $\psi_i = \psi \cup \{s_1, \dots, s_i\}$.

PROOF. Fix a sequence of states $Q \in \psi \rightarrow \psi'$ of length r . Then

$$\begin{aligned} f_{\text{avg}}(\psi') - f_{\text{avg}}(\psi) &= f_{\text{avg}}(\psi \cup \{s_1, s_2, \dots, s_r\}) - f_{\text{avg}}(\psi) \\ &= \mathbb{E} [f(\text{dom}(\psi \cup \{s_1, s_2, \dots, s_r\}), \Phi) - f(\text{dom}(\psi), \Phi) \mid \Phi \sim \psi'] \\ &= \mathbb{E} \left[\sum_{i=1}^r f_{s_i}(\text{dom}(\psi \cup \{s_1, \dots, s_{i-1}\}), \Phi) \mid \Phi \sim \psi' \right] \end{aligned}$$

where f_{s_i} is defined the conventional marginal gain function. Now, suppose s_i represents the (attempted) addition of an element e . Then

$$\begin{aligned} &\mathbb{E} [f_{s_i}(\text{dom}(\psi \cup \{s_1, \dots, s_{i-1}\}), \Phi)] \\ &= \mathbb{E} [f(\text{dom}(\psi \cup \{s_1, \dots, s_i\}), \Phi) - f(\text{dom}(\psi \cup \{s_1, \dots, s_{i-1}\}), \Phi)] \\ &= \mathbb{E} [f(\text{dom}(\psi \cup \{s_1, \dots, s_{i-1}\}) \cup \{e\}, \Phi) - f(\text{dom}(\psi \cup \{s_1, \dots, s_{i-1}\}), \Phi)] \\ &= \Delta(e \mid \psi \cup \{s_1, \dots, s_{i-1}\}) \end{aligned}$$

Then, by lemma B.1 and the linearity of expectation we can write

$$f_{\text{avg}}(\psi') - f_{\text{avg}}(\psi) = \sum_{t=1}^r \Gamma(j_t \mid \psi_{t-1}, \psi) \Delta(j_t \mid \psi)$$

If we then take the expectation with respect to Q , we get exactly the statement of the lemma. \square

THEOREM B.3. *For an adaptive monotone non-decreasing function f , the solution given by a greedy policy π selecting k elements satisfies:*

$$\left[1 + \left(\frac{f_{\text{avg}}(\pi^+)}{f_{\text{avg}}(\pi)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, \pi) + \epsilon_t) \right]^{-1} f_{\text{avg}}(\pi^*) \leq f_{\text{avg}}(\pi)$$

where π^+ is the greedy policy selecting $k+1$ elements, π^* is the optimal k -element policy, and $\hat{\Gamma}(i, S)$ is an estimator satisfying

$$\Gamma(x \mid \psi', \psi) \leq \hat{\Gamma}(i, \pi) + \epsilon_i$$

$$\forall \psi \in \pi(\emptyset), \psi \subset \psi', |dom(\psi') \setminus dom(\psi)| = i, x \notin dom(\psi') \cup dom(\psi)$$

PROOF. Suppose that we were to evaluate both the greedy and optimal policies to produce partial realizations ψ and ψ^* , respectively. Fix these partial realizations. Then, we have that $f(\psi^*) \leq f(\psi \cup \psi^*)$ by the definition of adaptive monotonicity and f non-decreasing. By Lemma B.2, we have:

$$f_{\text{avg}}(\psi^* \cup \psi) - f_{\text{avg}}(\psi) = \mathbb{E} \left[\sum_{t=1}^r \Gamma(j_t \mid \psi_{t-1}, \psi) \Delta(j_t, \psi) \mid Q \in \psi \rightarrow \psi \cup \psi^* \right]$$

As in the non-adaptive case, we make three relaxations. First, we make two relaxations to remove the dependence on knowledge of ψ^* : substituting $\hat{\Gamma}$ for Γ and g_{k+1} for j_t .

$$f_{\text{avg}}(\psi^*) - f_{\text{avg}}(\psi) \leq \mathbb{E} \left[\sum_{t=1}^r \left(\hat{\Gamma}(t-1, \psi) + \epsilon_{t-1} \right) \Delta(g_{k+1} \mid \psi) \right]$$

Next, we relax our knowledge of $|\psi^* \setminus \psi|$ from exact to none by replacing r with k . This also makes the content constant w.r.t. the expectation, and we therefore omit it.

$$f_{\text{avg}}(\psi^*) - f_{\text{avg}}(\psi) \leq \sum_{t=1}^k \left(\hat{\Gamma}(t-1, \psi) + \epsilon_{t-1} \right) \Delta(g_{k+1} \mid \psi)$$

Then, we rearrange terms to obtain

$$\left[1 + \left(\frac{\mathbb{E}[f_{\text{avg}}(\psi^*)]}{f_{\text{avg}}(\psi)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, \psi) + \epsilon_t) \right]^{-1} f_{\text{avg}}(\psi^*) \leq f_{\text{avg}}(\psi)$$

where ψ^+ is a ψ with g_{k+1} added. Then, taking the expectation w.r.t. the ψ and ψ^* values, we obtain

$$\left[1 + \left(\frac{f_{\text{avg}}(\pi^+)}{f_{\text{avg}}(\pi)} - 1 \right) \sum_{t=0}^{k-1} (\hat{\Gamma}(t, \pi) + \epsilon_t) \right]^{-1} f_{\text{avg}}(\pi^*) \leq f_{\text{avg}}(\pi)$$

with $\hat{\Gamma}(i, \pi) = \mathbb{E}[\Gamma(i, \psi) \mid \psi \in \pi(\emptyset)]$ the expected value with respect to partial realizations that could be produced by the policy π . \square

C PROPERTIES OF CURVATURE DEPENDENCY

The curvature-based definition of dependency has several interesting properties. While none are directly used in the above, they illuminate the structure ∇_f .

C.1 Curvature Dependency is Symmetric

LEMMA C.1. $u \in D(v) \iff v \in D(u)$

PROOF. First, we show that $v \notin D(u) \implies u \notin D(v)$. The proof is straightforward from the definition of primal curvature:

$$\begin{aligned} \nabla(u, v \mid T) = 1 &\implies f_u(T \cup \{v\}) = f_u(T) \\ &\implies f(T \cup \{u, v\}) - f(T \cup \{v\}) = f(T \cup \{u\}) - f(T) \\ &\implies f(T \cup \{u, v\}) - f(T \cup \{u\}) = f(T \cup \{v\}) - f(T) \\ &\implies \nabla(v, u \mid T) = 1 \end{aligned}$$

Next, we prove $v \in D(u) \implies u \in D(v)$ by contradiction. Assume that $v \in D(u)$ but $u \notin D(v)$. Thus, we have $\nabla(u, v \mid T) \neq 1$ for some choice of T , but $\nabla(v, u \mid T) = 1 \forall T$. By the above, we have that $\nabla(v, u \mid T) = 1 \implies \nabla(u, v \mid T) = 1$, which contradicts our assumption that $\nabla(u, v \mid T) \neq 1$. Thus, if $\nabla(u, v \mid T) \neq 1$ for some T , then we must also have that $\nabla(v, u \mid T) \neq 1$. Therefore, $v \in D(u) \implies u \in D(v)$. \square

From this result, and the observation that this relation is reflexive ($\nabla(x, x \mid S) = 0$ trivially unless $f_x(S) = 0$), we directly obtain that curvature dependency is a *dependency relation*.

C.2 Curvature Dependency Does Not Form a Matroid

LEMMA C.2. *The subset J of \mathcal{I} defined as $J = \{T \mid T \in \mathcal{I}, \nexists x, y \in T : x \in D(y)\}$ does not form a (sub-)matroid $\mathcal{M}' = (X, J)$.*

PROOF. The three conditions of a matroid $\mathcal{M} = (X, \mathcal{I})$ are as follows [21]:

- (1) $\emptyset \in \mathcal{I}$
- (2) If $I \in \mathcal{I}$ and $I' \subseteq I$, then $I' \in \mathcal{I}$.
- (3) If I_1 and I_2 are in \mathcal{I} and $|I_1| < |I_2|$, then there is an element $e \in I_2 \setminus I_1$ such that $I_1 \cup \{e\} \in \mathcal{I}$.

The first two conditions clearly hold for the pair (X, J) as defined in the statement of the lemma. We now show that the third does not by counterexample.

Suppose we have $M = (X, \mathcal{I})$, $X = \{a, b, c\}$, $\mathcal{I} = 2^X$, $J \subset \mathcal{I}$ as defined above. Now consider a scenario in which $D(a) = \{c\}$, $D(b) = \{c\}$ and $D(c) = \{a, b\}$. Note that this satisfies Lemma C.1. Then clearly $\{c\} \in J$ and $\{a, b\} \in J$. However, neither $\{a, c\} \in J$ nor $\{b, c\} \in J$. Therefore, condition (3) does not hold, implying $M' = (X, J)$ is not a matroid and therefore not a sub-matroid of M . \square

However, as properties (1) and (2) hold we see that this is an *independence system*.